

We claim:

1. A structural documentation system for converting a processing target electronic document described in a text format into a structured document having a predetermined document structure, said system comprising:

a reading module which reads definition information defining a correlation between elements as basic units configuring the document structure, and defining, for each of the elements, an extraction condition and an identifier thereof;

a retrieving module which refers to the extraction condition per element that is defined by the definition information read by said reading module, and which extracts a region coincident with the per-element extraction condition referred to out of the processing target electronic document; and

a structured document generating module which combines the regions extracted with respect to the respective elements by said retrieving module in accordance with the correlation between the elements that is defined by the definition information, and which generates the structured document by adding to each region an identifier defined by the definition information.

2. A structural documentation system according to claim 1, wherein said structured document generating module adds tags as an identifier in front and rear of each region extracted by

said retrieving module.

3. A structural documentation system according to claim
2, wherein said correlation between the elements defined by the
5 definition information takes a hierarchical structure in which
one element in a higher-order hierarchy embraces a plurality
of elements in a lower-order hierarchy,

said retrieving module extracts regions coincident with
respective extraction conditions of the elements in the
10 lower-order hierarchy out of a region extracted with reference
to an extraction condition of the element in its higher-order
hierarchy, and

said structured document generating module adds tags in
front and rear of the region extracted by said retrieving module
15 with respect to the element embracing no element in lower-order
hierarchy, and adds the tags for an element embracing elements
in lower-order hierarchy in front and rear of a region formed
by combining together the regions each extracted by said
retrieving module with respect to all the elements in the
20 lower-order hierarchy.

4. A structured documentation system according to claim
3, wherein said correlation between the elements shows a
hierarchical structure in which said element in a higher-order
25 hierarchy embraces an element in a lower-order hierarchy that
has a repetitive structure,

said retrieving module repeatedly extracts regions

coincident with the extraction condition of an element in the lower-order hierarchy having the repetitive structure out of the region extracted with reference to the extraction condition of the element in its higher-order hierarchy till no region
5 coincident with the extraction condition of the element in the lower-order hierarchy can be extracted, and

said structured document generating module adds common tags in front and rear of each of the regions extracted by said retrieving module with respect to the element in the lower-
10 order hierarchy.

5. A structural documentation system according to claim 3, wherein said correlation between the elements shows a hierarchical structure in which one element in a higher-order
15 hierarchy embraces a plurality of sequenced elements in a lower-order hierarchy and

said retrieving module extracts each region coincident with one of said extraction conditions of the elements in the lower-order hierarchy with reference to the extraction
20 condition of the sequenced element in the lower-order hierarchy out of a region from a portion just after an already-extracted region coincident with another extraction condition of the element in lower-order hierarchy within the region extracted with reference to the extraction condition of the element in
25 its higher order hierarchy.

6. A structural documentation system according to claim

1, wherein the extraction condition of any one of the elements defined by the definition information is a description pattern of the whole region to be extracted.

5 7. A structural documentation system according to claim 1, wherein the extraction condition of any one of the elements defined by the definition information is a description pattern of a start part of the region to be extracted and a description pattern of an end part thereof.

10

8. A structural documentation system according to claim 6 or 7, wherein the description pattern is expressed by a character string in the region to be extracted.

15

9. A structural documentation system according to claim 6 or 7, wherein the description pattern is expressed by a regular expression corresponding to the character string in the region to be extracted

20

10. A structural documentation system according to claim 1, wherein the extraction condition of any one of the elements defined by the definition information is a syntax element of the region to be extracted.

25

11. A computer readable medium stored with a program, executed by a computer to perform method comprising step of:
reading a processing target electronic document

described in a text format;

reading definition information which defines a correlation between elements as basic units configuring a document structure of a structured document, and which defines,
5 for each of the elements, an extraction condition and an identifier thereof;

referring to the extraction condition per element that is defined by the definition information read in said reading step;

10 extracting a region coincident with the per-element extraction condition referred to out of the processing target electronic document;

combining the regions extracted with respect to the respective elements in said extracting step in accordance with
15 the correlation between the respective elements that is defined by the definition information; and

generating the structured document by adding to each region an identifier defined by the definition information.

20190111 11:36:20